①

DTIC
ELECTE
S JUN 0 4 1990
D D

Technical Report ICMA-90-149   May 1990

On the Theory and Numerics of

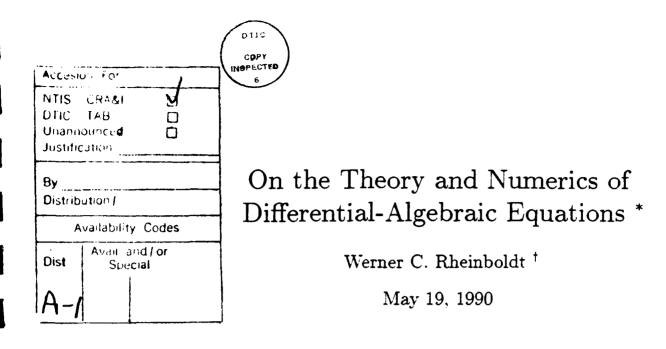# Differential-Algebraic Equations [1]

by

W. C. Rheinboldt
Institute for Computational Mathematics and Applications
Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15260

90 06 01 103

On the Theory and Numerics of Differential-Algebraic Equations *

Werner C. Rheinboldt [†]

May 19, 1990

1 Preface

This is a written version of a series of invited lectures on differential-algebraic systems of equations (DAEs) at the IVth SERC Numerical Analysis Summer School of Lancaster University. In line with the aims of the meeting these notes introduce some typical applications and basic properties of DAEs and then present an overview of recent, new existence theories for such systems based on differential geometric considerations and on a numerical approach derived from these theories. In the presentation the stress is on general concepts, results and applications rather than on detailed proofs.

Differential-algebraic systems of equations (DAEs) arise in many applications in science and engineering. For some examples we refer, for instance, to the monographs [3,22] and the many references given there. Three typical applications are sketched in Section 2 below. Over the years, it has become well known that the solution behavior of DAEs may differ considerably from that of standard ordinary differential equations (ODEs). A valuable measure of the deviation of a DAE from an ODE is the concept of an index which was first introduced in [21] and has since been formalized in various ways. The index highlights also some of the differences between the existence behavior of DAEs and ODEs although it does not, by itself, provide for any existence results. In fact, up to now, existence theories for nonlinear DAEs are available only for a few selected classes of systems.

Since the solutions of any DAE are expected to be smooth paths in some space of dependent variables, we should expect the equations to define a dynamical system in a suitable domain of that space. While this connection with dynamical systems

1

is immediately obvious for ODEs this is certainly not the case for DAEs and there appear to be only few studies that specifically address this connection (see e.g. [37] and [38]). Some aspects of the relationship between DAEs and dynamical systems will be discussed in Section 3.

In a series of papers [38], [41], and [36] a differential-geometric approach has been developed for the analysis of the dynamical system underlying a DAE and for the proof of general existence and uniqueness results for such systems. In Sections 4 and 5 the results in the two last-mentioned papers are summarized, moreover.Section 5 also addresses relations between some of these results and the index concept.

Comprehensive introductions to numerical methods for DAEs may be found in the cited monographs [3,22]. A brief survey of some of these methods is given in Section 6. Then in that section a new local parametrization approach for DAEs is presented which derives naturally from the differential-geometric existence theories and has been found to lead to very promising methods for the computational solution of higher-index DAEs. For the case of the Euler-Lagrange equations of constrained mechanical systems this approach includes the socalled method of generalized coordinate partitioning introduced first in [48].

2 Model Problems

This Section provides three illustrative examples of practical applications leading to differential-algebraic equations. As indicated before, there are numerous other areas were DAEs occur.

2.1 Constrained Dynamical Systems

A major source of DAEs is the kinematic and dynamic analysis of mechanical multibody systems. This is a venerable field of mechanics and we give here only some very simple examples and refer for further details to the extensive literature (see e.g. [23,49]).

Suppose that, under the influence of a force Q, a particle with mass m slides on a two-dimensional surface in R^3 specified by the real-valued equation $\Phi(x) = 0$. In order for the point to remain on the surface, a constraining force must act in the normal direction of the surface. If $D\Phi(x) \in L(R^3, R^1)$ denotes the derivative of Φ at any $x \in R^3$, then this normal direction is given by the vector $D\Phi(x)^T \in R^3$. Hence. by Newton's law we obtain here the DAE

$$(2.1) \qquad \Phi(x) = 0, \quad mx'' + zD\Phi(x)^T = Q$$

where $z \in R^1$ specifies the size of the constraining force. For example, suppose that the surface is a paraboloid and that gravity is the only force acting on the mass. then

(2.1) becomes

$$\begin{aligned}
x_1^2 - x_2^2 &= x_3 \\
mx_1'' - 2zx_1 &= 0 \\
mx_2'' - 2zx_2 &= 0 \\
mx_3'' - z &= mg.
\end{aligned}$$

(2.2)

More generally, suppose that the vector $x \in \mathbb{R}^n$ characterizes the configuration of all bodies of a mechanical system and that the kinematic constraints acting on the system are modelled by the s-dimensional (holonomic) constraint equations

$$\Phi(x, t) = 0. \tag{2.3}$$

Here Φ is now a mapping from $\mathbb{R}^n \times \mathbb{R}^1$ into $\mathbb{R}^s$, $1 \leq s < n$ and t represents time. Then the equations of motion are

$$M(x, t)x'' + D_x\Phi(x, t)^T z = Q(x, x', t) \tag{2.4}$$

where $M(x, t)$ is the mass matrix and $Q(x, x', t)$ the vector of applied forces.

As an example consider a simple, planar "slider crank" consisting of two bodies, namely, a bar of length 2 and a wheel of radius 1 centered at the origin of a (ξ, η) coordinate system in the plane. At one of its ends the bar pivots around a fixed point on the circumference of the wheel while its other end slides along the ξ-axis. Any configuration of the system may be characterized by the vector $x = (\alpha_1, \alpha_2, \xi)^T$ consisting of the current coordinate ξ of the bar's sliding end and the two angles α_1 and α_2 between the ξ-axis and the directions to either end of the bar, respectively. Thus, if the wheel turns with a constant torque τ, then the equations (2.4), (2.3) have here the form

$$\begin{aligned}
\cos\alpha_1 + 2\cos\alpha_2 &= \xi \\
\sin\alpha_1 - 2\sin\alpha_2 &= 0 \\
J_1\alpha_1'' - z_1\sin\alpha_1 + z_2\cos\alpha_1 &= \tau \\
m\xi'' - 2z_1\sin\alpha_2 - 2z_2\cos\alpha_2 &= 0 \\
J_2\alpha_2'' - z_1 &= 0
\end{aligned}$$

(2.5)

where J_1 and J_2 are the moments of inertia of the wheel and the bar, respectively, and m is the mass of the bar.

2.2 Electrical Circuits

A second extensive source of DAEs is the analysis of electrical circuits. Once again, we refer for details to the literature (see e.g. [10]) and discuss only the basic ideas and a simple example.

A circuit may be considered as an inter-connected collection of electrical devices. such as. resistors. inductors. capacitors, sources, etc. Its connection pattern is modelled by a finite. directed graph $\Omega = (N, \Lambda)$ with node set $N = \{1, 2, \ldots, p\}$ and branch set $\Lambda = \{\lambda_1, \ldots, \lambda_q\} \subset N \times N$ where single-node loops $(i, i) \in N \times N$ are excluded. Each branch corresponds to a specific component of the circuit. As an example consider a graph with four nodes with the following branches and components:

branch $(1, 2)$	linear resistor	resistance $= R_1$
branch $(2, 3)$	voltage source	voltage $= u_0$
branch $(1, 3)$	linear capacitor	capacitance $= C_1$
branch $(1, 4)$	linear inductor	inductance L_1
branch $(4, 1)$	linear capacitor	capacitance C_2
branch $(4, 3)$	linear capacitor	capacitance $= C_3$
branch $(3, 4)$	linear resistor	resistance $= R_2$.

Generally, the graph Ω can be characterized by its (node-arc) incidence matrix $A \in \mathbb{R}^{p \times q}$ with the elements

$$a_{ij} = \begin{cases} +1 & \text{if } (i, k) \in \Lambda \text{ for some } k \in N \\ -1 & \text{if } (k, i) \in \Lambda \text{ for some } k \in N \\ 0 & \text{otherwise.} \end{cases}$$

In our example the graph underlying the circuit then has the incidence matrix

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 & 1 & 1 & -1 \end{pmatrix}.$$

With each branch $\lambda_j = (i, k)$ of Ω two electrical quantities are associated. namely a current y_j and a voltage-drop u_j. They are connected by a functional relation

$$(2.6) \qquad \phi_j(y_j, u_j) = 0, \; j = 1, 2, \ldots, q,$$

the so-called *branch-characteristic* of λ_j. The specific form of (2.6) depends on the type of the device modelled by the branch, such as, for instance,

current source:	$y_j = \psi(t)$
voltage source:	$u_j = w(t)$
ideal diode:	$\max(u_j, -y_j) = 0$
linear resistor:	$u_j = Ry_j$
voltage driven resistor:	$y_j = \psi(u_j)$
current driven resistor:	$u_j = \psi(y_j)$
linear inductor:	$Cu_j' = y_j$
linear capacitor:	$Ly_j = u_j$

In our example, the set of branch characteristics (2.6) is given by the equations

$$
\begin{aligned}
R_1 y_1 &= u_1 \\
u_2 &= u_0 \\
C_1 u_3{}' &= y_3 \\
L y_4{}' &= u_4 \\
C_2 u_5{}' &= y_5 \\
C_3 u_6{}' &= y_6 \\
R_2 y_7 &= u_7.
\end{aligned}
$$

Kirchhoff's first conservation-law requires that the (algebraic) sum of the currents on the branches starting at a node must equal the sum of the currents on the branches terminating at that node. In terms of the incidence matrix A this means that a permissible current flow is characterized by any vector $y = (y_1, \ldots, y_q)^T \in \mathrm{R}^q$ for which

$$(2.7) \qquad\qquad A y = 0.$$

Kirchoff's second law specifies that the (algebraic) sum of all the voltage drops on the branches of any loop of Ω has to be zero. If we introduce the vector $u = (u_1, \ldots, u_q)^T \in \mathrm{R}^q$ of voltage drops, as well as the vector $w = (w_1, \ldots, w_p)^T \in \mathrm{R}^p$ of all nodal voltage levels, then the second law is corresponds to the equation

$$(2.8) \qquad\qquad u = A^T y , \; w_1 = 0$$

where the last equation was introduced to fix the absolute values of the nodal voltages.

Thus altogether (2.6),(2.7),(2.8) form a DAE of $2q + p - 1$ equations in $2q + p$ unknown. The reason for this difference is that A does not have full rank. If Ω is a connected graph - which is certainly a reasonable assumption - then a standard theorem of graph theory (see e.g. [6]) ensures that $\mathrm{rank} A = p - 1$. Thus one of the equations (2.7), is a linear combination of the others and hence may be dropped.

The equations (2.6), (2.7), (2.8) are called a descriptor form of the circuit. There are many ways of reducing the size of this system but we shall not enter into any details here.

2.3 Punch-Stretching of Sheet Metal

We end with a somewhat different example arising in connection with sheet metal stamping processes. Since in this case the formulation is somewhat more complex, we do not include all the details but refer instead to the literature (see e.g., [8,9]).

The processes to be considered involve the deformation of a sheet of metal in a forming press with a particular punch and die configuration. In order to ease the

discussion we consider a simpler problem. namely the so-called hydrostatic bulge test used widely in metallurgy. An initially flat sheet of metal is clamped over one end of a cylindrical chamber into which hydraulic oil is then pumped. This creates a hydrostatic load on the sheet and causes it to bulge outward.

In line with the formulation presented in [47] the equation of virtual work for the hydrostatic bulge deformation is given by

$$(2.9) \qquad h_0 \int_{A_0} \tau \delta\epsilon \ dA_0 = \int_{A_0} p\delta v \ dA_0.$$

Here h_0, A_0 denote initial sheet thickness and surface area. respectively. ξ is the radial distance to a material point on the sheet at time zero, and v is a volume measure. Moreover, $\tau = (\tau_1, \tau_2)$ is the vector of the Kirchhoff stresses in the radial and circumferential directions, respectively, while $\varepsilon = (\varepsilon_1, \varepsilon_2)$ is the vector of the logarithmic strains in the corresponding directions, defined by

$$\varepsilon_1 = \ln[(1 + u_\xi)^2 + w_\xi^2]^{1/2} \ , \quad \varepsilon_2 = \ln[1 + \frac{u}{\xi}].$$

where. u and w denote the radial and vertical displacements of the material point whose initial position is given by $(\xi, 0)$.

In addition, we have the material-dependent constitutive equations in rate form. These have the generic form

$$(2.10) \qquad \tau' = L\varepsilon' + r(\tau, \bar{\varepsilon}), \quad \bar{\varepsilon}' = g(\tau, \bar{\varepsilon})$$

where

$$L = \frac{E}{1 - \nu^2} \begin{pmatrix} 1 & \nu \\ \nu & 1 \end{pmatrix}$$

and E and ν are Young's modulus and Poisson's ratio, respectively, $\bar{\varepsilon}$ is the effective strain, and the nonlinear functions r and g depend on the specific form of the hardening law.

The volume measure in (2.9) is defined by

$$v = w(1 + \frac{u}{\xi})(1 + u_\xi)$$

and

$$V(t) = \int_{A_0} v(\xi) \ dA_0$$

is the volume of the bulge at time t which may be assumed, for instance, to satisfy $V(t) = \gamma t$ with fixed γ.

For the computation. we introduce finite element approximations of the displacements u, w, the stress components τ_1, τ_2, and the effective stress $\bar{\varepsilon}$. Then the equation (2.9) of virtual work is approximated by a nonlinear equation of the form

$$(2.11) \qquad F(x, y, q, t) = 0$$

where the vectors x, y, q contain the approximations of (u, w), (τ_1, τ_2), and p, respectively. Correspondingly. the constitutive equations (2.10) are approximated by a differential equation of the form

$$(2.12) \qquad \begin{aligned} y' - Lz' &= f_1(y, z) \\ z' &= f_2(y, z) \end{aligned}$$

where the vector z represents the approximation of the effective strain $\bar{\varepsilon}$. Thus. altogether the equations (2.11), (2.12) form a DAE.

3 DAEs and Dynamical Processes

3.1 DAEs and ODEs

The examples of Section 2 provide an indication of various possible forms of DAEs. In all cases the differential equations and algebraic equations turned out to be separated. Observe also that there may be variables for which no derivatives appear anywhere in the system. Of course, for the computation various specific properties of the equations are of particular interest. For instance, it is usually advantageous when the derivatives only occur linearly, etc.

In some applications the form of the equations may vary in different parts of the space. and, in particular, there may not exist a globally valid separation into algebraic equations and differential equations. Hence, in such cases, the DAEs have the generic form of an implicit differential equation

$$(3.1) \qquad F(x, x', t) = 0$$

which cannot be transformed into the form

$$(3.2) \qquad x' = f(x, t)$$

of an explicit ordinary differential equation (ODE)

If, in (3.1), F is a sufficiently smooth map from R^{2n+1} into R^n and the derivative $D_p F(x, p, t) \in L(R^n, R^n)$ is an isomorphism at some solution (x, y, t) of $F(x, y, t) = 0$ then the implicit function theorem guarantees that (3.1) can be transformed locally

into the form (3.2). Hence, in our setting we should assume that $D_p F(x, p, t)$ does not have full rank. More specifically, we shall call (3.1) an implicit DAE only if

$$(3.3) \qquad \operatorname{rank} D_p F(x, p, t) = \text{constant} < n,$$

on the domain under consideration. This constant-rank assumption excludes various singular implicit equations (3.1) with a solution behavior that may differ radically from that of ODEs or DAEs (see e.g. [35]).

The existence and uniqueness theory for solutions of explicit ODEs (3.2) is a well-developed subject (see e.g. [11] and also the Appendix). In particular, if $f : E \subset R^{n-1} \to R^n$ is of class C^1 on some open set E, then we know that for any point $(x_0, t_0) \in E$ there exists a C^2-solution $x : J \subset R^1 \to E$ of (3.2), defined on some open interval J containing t_0, which satisfies the initial condition $x(t_0) = x_0$. Moreover, any two such solutions satisfying the same initial condition are identical on the intersection of their domain.

This local solvability result for ODEs does not carry over directly to DAEs. In fact, consider the simple system

$$(3.4) \qquad \begin{aligned} x_1 &= \cos x_2 \\ x_1' &= x_3 \\ x_2' &= 1, \end{aligned}$$

and suppose that $x : J \subset R^1 \to R^3$ is any C^1-solution of (3.4) on some open interval J. Then, by differentiating $x_1(t) - \cos x_2(t) = 0$ with respect to t and using the differential equations, we find that the solution must satisfy the algebraic condition

$$(3.5) \qquad x_3 + \sin x_2 = 0,$$

whence necessarily

$$(3.6) \qquad x(t) = \begin{pmatrix} \cos t \\ t \\ -\sin t \end{pmatrix}.$$

Conversely, (3.6) does define a C^∞-solution $x : R^1 \to R^3$ of (3.4). In other words, (3.6) is the only solution of (3.4) and we certainly cannot prescribe any initial conditions other than, trivially, a point of (3.6).

This result indicates that for DAEs there may be some "hidden" constraints, such as (3.5), which all solutions have to satisfy. As a consequence, there may not exist any solution through every choice of initial point; that is, only certain initial conditions may be admissible.

Obviously, any "hidden" constraints may be expected to cause difficulties during the numerical solution of the DAE. This is indeed the case, and, in fact, it is by

now well known that the degree of difficulty rises with the number of such additional constraints. This observation led W.Gear and L. Petzold ([21]) to introduce an index which measures the "deviation" of a DAE from an ODE. We shall discuss later some aspects of this concept; at this moment it will suffice to characterize the index of a DAE loosely as the total number of given algebraic and hidden constraints that are needed to specify the solution completely. In this sense, (3.4) is a DAE with index 2.

As another example consider the DAE (2.2) modelling the dynamics of a mass-point on a paraboloid. This system can be written in the first order form

$$
(3.7) \qquad
\begin{aligned}
\Phi(x) &\equiv x_1^2 - x_2^2 - x_3 = 0 \\
x' &= y \\
y' &= ge^3 - zD\Phi(x)^T
\end{aligned}
$$

where $x, y \in \mathbb{R}^3$, e^3 is the third natural basis vector of $\mathbb{R}^3$ and we replaced z/m by z. By differentiating the algebraic equation and using the differential equations we obtain as first "hidden" constraint

$$
(3.8) \qquad 2x_1y_1 - 2x_2y_2 - y_3 = 0.
$$

In turn, by differentiating (3.8) we are led to the further constraint

$$
(3.9) \qquad 2(y_1^2 + y_2^2) - z(1 + 4(x_1^2 - x_2^2)) = 0.
$$

It is not difficult to see that (3.7) together with these two constraints (3.8), (3.9) completely specifies the solutions. Thus in our terminology the DAE has index 3.

The two constraints have here a simple geometric meaning. In fact, (3.8) requires the velocity vector y to be tangential to the paraboloid while (3.9) means that the constraining force $zD\Phi(x)^T$ has to balance the other two forces.

This index-result is not restricted to the special example (2.2); in fact it turns out that all Euler-Lagrange systems (2.3), (2.4) have index 3.

3.2 Dynamical Processes

Ordinary differential equations are a fundamental tool in the study of dynamical processes that are finite-dimensional, differentiable and causal. By this we mean processes for which

(i) the states are characterized by finitely many degrees of freedom,

(ii) the changes of the states are described by differentiable functions, and

(iii) the future behavior is uniquely determined by the initial conditions.

Our examples suggest that differential-algebraic equations also represent models of such dynamical processes.

The theory of dynamical processes has been heavily influenced by mechanical considerations. Thus. we use a simple mechanical example to review some of the basic terminology. Consider the motion of k particles in R^3 and let x_i and y_i. $i = 1.2.\ldots.k$. denote the location and velocity of the particles at time t. Then with $q = (x_1,\ldots,x_k) \in R^{3k}$ and $p = (y_1,\ldots,y_k) \in R^{3k}$ the state of the system is given by $(q.p)$. The states are usually restricted to some specified subset $S \subset R^{3k} \times R^{3k}$ - the state space of the system.

In many cases. the state space is an open subset of $R^{3k} \times R^{3k}$. For instance. if no two particles are ever allowed to be in the same place at the same time. then

$$S = \{(q.p) \in R^{3k} \times R^{3k}: x_i \neq x_j \text{ for } i \neq j. \ i,j = 1.2,\ldots.k\}$$

is certainly open.

On the other hand - for instance when there are angular variables or constraints - the state space need not be an open set. For example. consider a system of rigidly connected particles. Then the configuration vectors q have to belong to the set

$$C = \{q \in R^{3k}; \| x_i - x_j \|_2 = c_{ij}: \text{ for } i \neq j. \ i,j = 1.2,\ldots,k\}$$

where c_{ij} are given constants. For $k \geq 4$ this is a six-dimensional submanifold [1] of R^{3k}. This follows from the well-known fact that the position of a rigid body in R^3 is uniquely characterized by the location of one point and the orientation of an orthonormal coordinate system fixed within the body. Then the state space may be identified with the tangent bundle $C \times R^{3k}$ of C and hence is a 12-dimensional submanifold of $R^{3k} \times R^{3k}$.

In differential-geometric terms the two cases are not very different. In fact. any open subset of $R^{3k} \times R^{3k}$ is a $6k$-dimensional submanifold of that space and hence. in either case, the state space is a submanifold. This agrees with the fundamental assumption introduced by H. Poincarè ($\sim$ 1880) that the state space of a mechanical system should be a differentiable manifold. Correspondingly, the dynamical system is viewed as a field of vectors on this manifold such that a solution is a smooth curve tangent at each of its points to the vector attached to that point. We refer. e.g., to [1] and the historical references included there.

However, from a computational viewpoint. there is indeed a substantial difference between the above two cases which, in fact, reflects again the earlier indicated differences between ODEs and DAEs. As noted before, in the classical theory of explicit ODEs (3.2) we assume f to be of class C^1 on some *open* subset E of R^{n+1} and, of course, on E the ODE induces the natural vector field

$$(x,t) \in E \quad \longmapsto \quad ((x,t), f(x,t)) \in TE = E \times R^{n+1}.$$

---

[1] Some differential-geometric concepts and results are collected in the Appendix.

Since the points (x,t) represent the states of the system. this corresponds to the case when the state space is an open subset. On the other hand. as the above example of rigidly connected particles shows. for a DAE the state space is expected to be a lower dimensional manifold. For instance. in the trivial example (3.4) this is the one-dimensional submanifold

$$\{x \in \mathrm{R}^3;\ x_1 - \cos x_2 = 0,\ x_3 + \sin x_2 = 0\}$$

defined by the given and hidden constraints.

In the standard theory of numerical methods for solving ODEs of the form (3.2) it is critical that the domain E of the right-hand side is open and that there exists a locally unique solution of (3.2) through each point x of E. In fact. any such method generates a sequence $\{(x_k, t_k); k = 1, 2, \ldots\}$ of approximating points on the solution through a given initial point (x_0, t_0). At best, we know that these points belong to some open neighborhood of the exact solution contained in the open set E . Furthermore. for the step from x_k to x_{k-1} all methods are designed to approximate the *local* solution through x_k in the sense that the error between x_{k+1} and this local solution converges to zero when the steplength tends to zero. It is by no means obvious how to extend this approach to the case when the domain E is no longer an open set but some lower-dimensional submanifold of $\mathrm{R}^n \times \mathrm{R}^1$.

4 Existence Theory for Implicit DAEs

As noted earlier, the literature on DAEs is growing rapidly but general existence theories have only begun to be developed relatively recently. The earliest such result appears to be the existence theory for gradient systems

$$\nabla_x g(x,y) = 0 ,\ \ x' = f(x,y).$$

developed by F.Takens [45] who used the approximating. singularly perturbed system of differential equations

$$\varepsilon x' = \nabla_x g(x,y) ,\ \ x' = f(x,y).$$

with asymptotically small ε.

Best understood are probably the linear DAEs with constant coefficients

$$(4.1) \qquad Ax' + Bx = g(t),\ x \in \mathrm{R}^n,\ A, B \in L(\mathrm{R}^n),\ \mathrm{rank} A = r < n$$

for which existence results can be proved by means of the Kronecker canonical form for matrix pencils (see, e.g., [16]). For a presentation of this theory see e.g. [21] or [22]. For further references see also [26,46].

In [38] the indicated interpretation of DAEs as dynamical systems on manifolds was used to obtain existence results for semi-explicit systems

$$F_1(x) = 0, \quad A(x)x' = G(x).$$

These results were generalized in [41] to first and second order systems of the form

$$(4.2) \qquad\qquad F_1(x) = 0, \quad F_2(x, x', z) = 0,$$

and
$$(4.3) \qquad\qquad F_1(x) = 0, \quad F_2(x, x', x'', z) = 0.$$

respectively.

For general implicit equations (3.1) – of course, under the constant-rank assumption (3.3) – local existence results were first given in [22], however, under the restrictive condition that $\ker D_p F(x, p, t)$ is independent of x and p. Finally, without such a condition, a solution theory for implicit DAEs was presented in [36]. This theory will be outlined in the following sub-section: for proofs and further details we refer to the original article.

4.1 Local Theory for Implicit DAEs

For ease of notation we shall consider (3.1) in the autonomous form

$$(4.4) \qquad\qquad F(x, x') = 0,$$

under the three assumptions

(A1) $F : E \subset \mathrm{R}^n \times \mathrm{R}^n \to \mathrm{R}^n$ is C^2 on the open set E;

(A2) $\mathrm{rank} DF(x, p) = n, \ \forall \ (x, p) \in E$;

(A3) $\mathrm{rank} D_p F(x, p) = r < n, \ \forall \ (x, p) \in E$.

The condition (A2) requires that the equations (4.4) are independent and also implies that $F^{-1}(0)$ is an n-dimensional C^2-submanifold of $\mathrm{R}^n \times \mathrm{R}^n$ (see again the Appendix).

An instructive prototype for (4.4) is the semi-implicit DAE

$$(4.5) \qquad\qquad F(x, x') = \left(\begin{array}{c} F_1(x) \\ F_2(x, x') \end{array} \right)$$

Here (A1) holds if $F_1 : E_x \to \mathrm{R}^{n-r}$ and $F_2 : E_x \times E_p \to \mathrm{R}^r$, are of class C^2 on open sets E_x and $E = E_x \cdot E_p$, respectively. Moreover, (A2) and (A3) are satisfied if

$\mathrm{rank} DF_1(x) = n - r,\ \forall x \in E_x$ and $\mathrm{rank} D_p F_2(x,p) = r,\ \forall (x,p) \in E$. Obviously, the first of these conditions implies that

$$(4.6) \qquad\qquad M = \{x \in E_x; F_1(x) = 0\}$$

is an r-dimensional C^2-submanifold of R^n.

A C^k-solution of the general equation (4.4) is any mapping

$$x : J \ - \ R^n,\ (x(t), x'(t)) \in E,\ F(x(t), x'(t)) = 0,\ \forall t \in J$$

that is of class C^k on some open interval J of R^1. As noted in the previous Section, we cannot expect that there is a solution through each point $(x,p) \in E$ and the following lemma provides a necessary condition for this to hold:

Lemma 4.1 *For* $(x,p) \in E$ *the conditions*

$$(4.7) \qquad\qquad F(x,p) = 0,\ D_x F(x,p)p \in \mathrm{rge} D_p F(x,p),$$

are neccessary for the existence of a C^1*-solution of (4.4) that passes through* (x,p).

For any C^2-solution of (4.4) this follows directly from the fact that by differentiation of $F(x(t), x'(t)) = 0$ we obtain $D_x F(x(t), x'(t))x'(t) - D_p F(x(t), x'(t))x''(t) = 0$ for all t in J. Of course, for C^1-solutions this argument cannot be used and a more subtle proof is required (see [36]).

For a closer analysis of the set of points characterized by the necessary conditions (4.7) we introduce the orthogonal projections

$$
\begin{aligned}
P, Q : E \to L(\mathrm{R}^n, \mathrm{R}^n),\ P(x,p)\mathrm{R}^n\ &=\ \mathrm{rge} D_p F(x,p), \\
Q(x,p)\ &=\ I_n - P(x,p),\ \forall (x,p) \in E,
\end{aligned}
$$

Because of the constant rank condition (A3) these projections are C^1-functions on E. Hence, also the reduced map

$$(4.8) \quad \hat{F} : E \to \mathrm{R}^n, \hat{F}(x,p) = P(x,p)F(x,p) + Q(x,p)D_x F(x,p)p,\ (x,p) \in E,$$

is of class C^1 on E. Then we can show that the set E_N of all points satisfying the necessary conditions (4.7) is given by

$$(4.9) \qquad\qquad E_N = \{(x,p) \in E;\ F(x,p) = 0, \hat{F}(x,p) = 0\}.$$

In the special case of (4.5) the projections are independent of x and p. In fact we have

$$P = \begin{pmatrix} 0_{n-r} & 0 \\ 0 & I_r \end{pmatrix}$$

and therefore

(4.10) $$\hat{F}(x,p) = \begin{pmatrix} DF_1(x)p \\ F_2(x,p) \end{pmatrix}$$

and

$$E_N = \{(x,p) \in E; \ F_1(x) = 0, \ DF_1(x)p = 0, \ F_2(x,p) = 0 \ \}.$$

Generally, the mapping (4.8) defines the reduced equation

(4.11) $$\hat{F}(x,x') = 0,$$

which, in essence, has the same solutions has the original DAE. More specifically the following result holds:

Lemma 4.2 *Any C^1-solution of the original equations (4.4) solves the reduced equation (4.11). Conversely, any C^2-solution of (4.11) that passes through some point of $F^{-1}(0)$ is a C^2-solution of (4.4).*

In the special case of (4.5) when F_2 is linear in p; that is, in the case of the DAE

(4.12) $$F(x,x') = \begin{pmatrix} F_1(x) \\ A(x)x' - G(x) \end{pmatrix} = 0,$$

$D_p\hat{F}(x,p)$ does not depend on p and the reduced equation (4.11) becomes the linear equation

(4.13) $$B(x)x' = \begin{pmatrix} 0 \\ G(x) \end{pmatrix} = 0, \quad B(x) = \begin{pmatrix} DF_1(x) \\ A(x) \end{pmatrix}.$$

Suppose that the subset

(4.14) $$M_0 = \{x \in M; B(x) \in \mathrm{Isom}(R^n)\}$$

of the constraint manifold (4.6) is not empty. Then M_0 is an r-dimensional submanifold of R^n on which (4.13) induces the tangential C^1-vectorfield

$$x \in M_0 \ \longmapsto \ (x,p) \in TM_0, \quad p = B(x)^{-1}\begin{pmatrix} 0 \\ g(x) \end{pmatrix}.$$

It is readily seen that the integral curves of this vectorfield are exactly the solutions of (4.13) and therefore, by Lemma 4.2, also of the DAE (4.12). This corresponds to the approach used in [38] to develop an existence theory for DAEs of the form (4.12).

For the general DAE (4.4), we proceed analogously and assume that the set

(4.15) $$E_A = \{(x,p) \in E_N; \ D_p\hat{F}(x,p) \in \mathrm{Isom}(R^n)\}$$

is non empty. Clearly, by continuity E_A is (relatively) open in E_N. Moreover, by the implicit function theorem it follows that locally in some open neighborhood of any point $(x_0, p_0) \in E_A$ the reduced equation (4.11) can be transformed into an explicit ODE. Thus by applying the standard ODE-theory we can now prove the following local existence result for the original DAE (4.4):

Theorem 4.1 *Given $t_0 \in R^n$, consider the initial value problem*

$$(4.16) \qquad\qquad F(x, x') = 0, \quad x(t_0) = x_0, \quad x'(t_0) = p_0$$

under the assumptions (A1.2.3). If a C^1-solution of (4.16) exists, then $(x_0, p_0) \in E_N$. Conversely, for any $(x_0, p_0) \in E_A$ there exists a C^1-solution of (4.16) which is unique on some sufficiently small interval J containing t_0. Moreover, this solution is actually of class C^2 on J.

The set E_A of (4.15) has a manifold structure. This is self-evident in the case of (4.5) where, obviously, the derivative of the mapping

$$(x, p) \in E \quad \longrightarrow \quad \begin{pmatrix} F_1(x) \\ \hat{F}(x, p) \end{pmatrix} \in R^{2n-r}$$

has full rank for any point of E_A and hence, E_A is either empty or an r-dimensional C^1-submanifold of $R^n \times R^n$. But the result also holds in general:

Lemma 4.3 *The set $E_A \subset E$ of admissible initial points of (4.4) is either empty or an r-dimensional C^1-submanifold of $R^n \times R^n$.*

From the implicit function theorem it follows that for any point $(x_0, p_0) \in E_A$ there exist an open neighborhood $U = S_x \times S_y \subset E$ and a unique C^1-mapping $\eta : S_x \longrightarrow S_y$ with $\eta(x_0) = p_0$ such that $\hat{F}(x, p) = 0$ for $(x, p) \in U$ if and only if $p = \eta(x)$. Since E_A is open in E_N we may assume that $U_0 = U \cap E_A = U \cap E_N$. Let $\Pi : R^n \times R^n \longrightarrow R^n$ be the projection onto the first factor. Then, the result means that the restriction ΠU_0 is a C^1-diffeomorphism from U_0 onto ΠU_0. Hence

$$M = S_x \cap \Pi U_0$$

is an r-dimensional submanifold of R^n and

$$x \in M \quad \longmapsto \quad (x, p) \in TM, \quad p = \eta(x),$$

is a tangential C^1-vectorfield on M for which it can be shown that the integral curves are exactly the solutions of (4.4) in E_0. In other words, the following result holds:

Theorem 4.2 *With the above terminology, $x : J \to \mathrm{R}^n$ is a C^1-solution of $F(x, x') = 0$ satisfying $(x(t), x'(t)) \in U$ for $t \in J$ if and only if $x(t) \in M$, $x'(t) = \eta(x(t))$, $\forall t \in J$.*

Thus, as expected, the DAE (4.4) is locally equivalent to an explicit ODE on an r-dimensional submanifold of R^n.

The results in this section can be extended easily to the general nonautonomous case
$$F(t, x, x') = 0.$$
In fact, as usual we can transform this problem into autonomous form by introducing the mapping

$$G : \mathrm{R}^{n+1} \times \mathrm{R}^{n+1} \to \mathrm{R}^{n+1}, \quad G((t, x), (\tau, p)) = \begin{pmatrix} \tau - 1 \\ F(t, x, p) \end{pmatrix}.$$

Then under the required smoothness assumptions the necessary conditions (4.7) for G assume the form

$$F(t, x, p) = 0, \quad D_t F(t, x, p) + D_x F(t, x, p) p \in rge\, D_p F(t, x, p).$$

Similarly, we can derive the form of the reduced mapping and of the set E_A of admissible initial points (see [36]).

4.2   Globalizations

Consider again the implicit DAE (4.4) under the assumptions (A1.2.3). As we saw in Theorem 4.1, for any (x_0, p_0) in the set E_A of (4.15) the initial value problem (4.16) has a local C^1-solution on a sufficiently small open interval $J = (a, b) \in \mathrm{R}^n$ containing t_0. As in the standard ODE-theory, under appropriate conditions these local solutions can be continued.

In [36] the following basic continuation result was proved:

Theorem 4.3 *Suppose that $E_A = E_N$.*

(i) *If for some $\epsilon \in (0, b - a)$ the set $\{p \in \mathrm{R}^n; p = x'(t), b - \epsilon < t < b\}$ is bounded then $\lim_{x \to b-} x(t) = x_b$ exists.*

(ii) *If $\lim_{t \to b-} x(t) = x_b$ exists and for some sequence $\{t_k\} \in J$ with $\lim_{k \to \infty} t_k = b$, the sequence $\{x'(t_k)\}$ has an accumulation point p^* for which $(x_b, p^*) \in E$ then $\lim_{t \to b-} x'(t) = p^*$ and hence, for $b < \infty$ the solution can be continued to the right.*

An analogous result holds for the left endpoint. Thus any local solution of (4.16) can be extended to a maximal open interval $J^* = (a^*, b^*)$, $-\infty \leq a^* < b^* \leq \infty$.

For an explicit initial value problem

$$x' = f(x), \quad x(0) = x_0, \quad x \in \mathrm{R}^n$$

with smooth f on all of R^n, any maximally extended solution for which $\lim_{t \to b_-} x(t)$ exists has bounded derivatives $x'(t)$ near b. This does not carry over to DAEs as the following result shows (see again [36]):

Theorem 4.4 *Suppose that $E = \mathrm{R}^n$ and $E_A = E_N$. If $b^* < \infty$, then $x'(t)$ is unbounded on the interval $(b^* - \epsilon, b^*) \in J^*$ for all sufficiently small $\epsilon > 0$. Hence, if $\lim_{t \to b_-} x(t) = x_b$ exists, then $\lim_{t \to b_-} \|x'(t)\| = \infty$.*

Again, an analogous result holds for the left endpoint.

Theorem 4.2 showed that the implicit DAE (4.4) is locally equivalent to a vector field on some r-dimensional submanifold of R^n. These local vectorfields can be extended by applying the theory of covering spaces. This was first used in [41] in connection with the DAEs (4.2) and (4.3) and then extended to the implicit case in [36].

We sketch only briefly the general approach. Clearly, the local result shows that the restriction $\Pi_! E_A$ is a local homeomorphism between E_A and ΠE_A. Let E_A^* be some non-empty, arc-connected subset of E_A for which (E_A^*, Π_A^*), with $\Pi_A^* = \Pi E_A^*$, is a covering space of ΠE_A^*. In other words, each point $x \in \Pi E_A^*$ is assumed to have an open, arc-connected neighborhood U such that each arc-component of $(\Pi_A^*)^{-1} U$ is not empty and is mapped topologically onto U by ΠE_A^*. Often $E_A^* = E_A$ can be used here. This is certainly the case when for fixed $x \in \Pi E_A$ there are only finitely many p with $(x, p) \in E_A$. For instance, this holds for the semi-linear DAE (4.5). In general, it is always possible to choose E_A^* as the closure of a non-empty, pre-compact, (relatively) open, and arc-connected submanifold of E_A.

For any given $(x_0, p_0) \in E_A^*$ let now M^* be a non-empty, (relatively) open, simply connected subset of ΠE_A^* that contains x_0. For any $x \in M^*$ choose a path $\xi : J \to M^*$ which connects x_0 with x. Then there exists a unique lifting $\xi^* : J \to E_A$ with initial point (x_0, p_0) for which $\Pi_A^* \xi^* = \xi$. This lifted path has a unique endpoint (x, p) in E_A because all paths in M^* between x_0 and x are homotopic. Since x was arbitrary in M^* our local result can now be used to prove that M^* indeed is an r-dimensional submanifold of R^n and that the DAE (4.4) induces a tangential vector field on M^* for which all integral curves in M^* are solutions of (4.4).

5 DAEs with Higher Index

5.1 Linear, Constant Coefficient DAEs

As mentioned before, the linear problems with constant coefficients (4.1) probably represent the most extensively studied DAEs in the literature. For sufficiently smooth g the necessary condition (4.7) turns out to have the form $Bx - g'(t) \in \text{rge}A$ and it is easily seen that $E_A \neq \emptyset$ exactly if

(5.1) $\qquad\qquad Au = 0 \quad \text{and} \quad Bu \in \text{rge}A \quad \text{imply} \quad u = 0.$

The cited existence theory for (4.1) ensures solvability if and only if the matrix pencil (A, B) is regular; that is, if there is some $\lambda \in \mathbb{R}^1$ such that $B - \lambda A$ is invertible. A central concept in the solvability theory of (4.1) is its index which is defined to be the index of the coefficient pencil assumed to be regular. For any regular pencil (A, B) let λ be such that $B - \lambda A \in \text{Isom}(\mathbb{R}^n)$, then the index of the pencil is the smallest integer κ such that

$$\ker\left[(B - \lambda A)^{-1}A\right]^{\kappa+1} = \ker\left[(B - \lambda A)^{-1}A\right]^{\kappa}$$

It can be shown that κ is finite and independent of the choice of λ, (see e.g. [22]), and it is also readily seen that $\kappa = 0$ if and only of A is invertible; that is, if (4.1) is equivalent with an explicit ODE.

In order to relate the theory of Section 4 to this index-concept, let $P \in L(\mathbb{R}^n)$ be the orthogonal projection onto rge$A$ and set $Q = I_n - P$. As before we differentiate the DAE (4.1) and then multiply the resulting equation $Ax'' - Bx' = g'(t)$ by Q in order to remove again the second derivative of x. Together with the projection of the original equation onto rge$A$ this produces the reduced DAE

(5.2) $\quad A_1 x' - B_1 x = g_1(t), \quad A_1 = PA + QB, \quad B_1 = PB, \quad g_1(t) = Pg(t) + Qg'(t).$

Even without recourse to the earlier theory, it is readily checked that (5.1) is equivalent with $A_1 \in \text{Isom}(\mathbb{R}^n)$ and, hence, that when (5.1) holds then (5.2) can be transformed into an explicit ODE.

Suppose therefore that A_1 is singular. Then we may apply the same procedure repeatedly, as often as necessary, to obtain a sequence of DAEs of the form

(5.3) $\qquad\qquad A_j x' + B_j x = g_j'(t), \quad j = 0, 1, \ldots,$

where A_j, B_j, g_j are specified recursively by $A_0 = A$, $B_0 = B$, $g_0 = f$ and

(5.4) $\qquad\qquad \begin{aligned} A_{j+1} &= P_j A_j + Q_j B_j, \quad B_{j+1} = P_j B_j, \\ g_{j+1}(t) &= P_j g_j(t) + Q_j g_j'(t), \quad j = 0, 1, \ldots. \end{aligned}$

while P_j denote the orthogonal projections onto $\mathrm{rge}A_j$ and $Q_j = I_n - P_j$. The process stops with the smallest integer k such that A_k is invertible.

The following result, proved in [36], shows that this integer k is exactly the index of the DAE:

Theorem 5.1 *If the matrix pencil (A, B) is regular and* $\mathrm{rank}A < n$ *(so that* $\kappa \geq 1$*) then* $k = \kappa < \infty$. *Conversely, if* $k < \infty$ *then* (A, B) *is regular and* $k = \kappa$.

In [36] it was also shown that the theory of Section 4 provides all the solutions of (4.1) provided only that g is smooth enough.

5.2   Nonlinear Problems with Higher Index

The discussion of the previous section suggests that we may proceed analogously when the set E_A is empty for the general implicit initial value problem

(5.5) $$F(x, x') = 0, \quad x(0) = x_0, \quad x'(0) = p_0.$$

The first step in the construction of a sequence of problems corresponding to (5.3), (5.4) was already done in Section 4. In fact, we differentiated the DAE and then applied the projections P and Q to obtain the reduced equation (4.11).

Our sufficient condition is that $D_p\hat{F}(x, p)$ is invertible at the given initial point $(x_0, p_0) \in E_N$. If this sufficient condition does not hold, then, as in the linear case, it is natural to construct recursively the sequence of mappings

(5.6) $$\begin{aligned} F^0 &= F, \quad F^1 = \hat{F}, \\ F^{j+1} &= P_j(x, p)F^j(x, p) + Q_j(x, p)D_xF^j(x, p), \quad j = 0, 1, \ldots \end{aligned}$$

where P_j again is the orthogonal projection onto D_pF^j and $Q_j = I_n - P_j$. The process is repeated until D_pF^k is invertible at the point under consideration.

As before, one might consider calling this integer k the local index of the problem at the particular point. However, the situation differs here in a critical aspect from that of the linear case. The theory of Section 4 can be applied to the map F^k only if the conditions (A1,2,3) are valid for all the maps (5.6) in some neighborhood E_0 of the point (x_0, p_0). In particular, we require $\mathrm{rank}D_pF^j(x, p)$ to be constant in such a neighborhood for the projections P_j, Q_j and hence F^j to be of class C^1. In addition, we also need the three conditions to conclude from the non-singularity of D_pF^k that the system $F^k(x, x') = 0$ can be tranformed locally into an explicit ODE and that the original problem (5.5) has a unique solution. As mentioned earlier, the existence theory for (5.5) changes considerably when the constant-rank condition is violated (see [35]).

In line with this, the problem (5.5) will be defined to have local index k at (x_0, p_0) if there is some open neighborhood E_0 of that point such that for $j = 0, 1, \ldots, k - 1$

the mappings (5.6) satisfy the conditions (A1.2.3) and $D_p F^k(x_0, p_0)$ is invertible. Note that this index does not merely depend on information at the given point and hence has a global nature. Obviously, the theory developed in Section 4 assumes that the implicit problem (4.4) has index one.

5.3 Semi-Implicit Problems with Higher Index

The recursive analysis outlined in the previous section is a powerful tool for the theoretical study of higher index problems. But for specific classes of equations it is often easier to derive existence and uniqueness results directly. As an example of this we consider in this section the special problems (4.2) and (4.3) both of which have, in general, index higher than one.

More specifically, for the system

$$(5.7) \qquad\qquad F(x, x', z) = \left(\begin{array}{c} F_1(x) \\ F_2(x, x', z) \end{array} \right)$$

suppose that

(B1) $F_1 : E_x \in \mathbb{R}^n \longrightarrow \mathbb{R}^s$ is C^2,

(B2) $F_2 : E_2 = E_x \times E_p \times E_z \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^r$ is C^1,

$$(B3) \quad \left(\begin{array}{cc} DF_1(x) & 0 \\ D_p F_2(x, p, z) & D_z F_2(x, p, z) \end{array} \right) \in \mathrm{Isom}(\mathbb{R}^{m-n}), \quad \forall (x, p, z) \in E_2$$

where $s \leq n \leq s + r = n + m$ and $E_x, E_p \subset \mathbb{R}^n, E_z \subset \mathbb{R}^m$ are non-empty, open sets.

Evidently, these conditions imply that (A1.2,3) are satisfied for (5.7). The assumption (B3) is often called the index-two condition since the index of (5.7) can be at most two. Of course, in the degenerate case $s = m = 0$, (B3) implies that (5.7) can be transformed into an explicit ODE and thus has index zero. Moreover, it is easily seen that for $s = 0$, $m \neq 0$ or $m = 0$, $s \neq 0$ the index is one.

From (B3) it follows that $\mathrm{rank} DF_1(x) = s$ whence each member of the family of sets

$$(5.8) \qquad\qquad M_b = \{x \in E_x; \ F_1(x) = b\}, \quad b \in F_1(E_x)$$

is an $(n - s)$-dimensional C^2-submanifold of $\mathbb{R}^n$. Evidently, any $x_0 \in E_x$ belongs to the unique constraint manifold (5.8) specified by $b = F_1(x_0)$. Moreover, for a solution of (5.7) to pass through this point we must have $x_0 \in M_0$.

Let $x : J \in E_x$ be any C^1-path defined on some open interval $J \in \mathbb{R}^1$. If x is a p on M_0; that is, if $F_1(x(t)) = 0$, $\forall t \in J$, then necessarily

$$(5.9) \qquad\qquad DF_1(x(t))x'(t) = 0, \quad \forall t \in J;$$

that is. $t \in J - (x(t), x'(t))$ has to be a path on the tangent bundle of M_0. The explicit use of the differentiated constraint equation $DF_1(x)x' = 0$ is a basic step in the so-called index-reduction technique for rewriting the DAE as a lower index system (see [18]). But (5.9) can also be interpreted in another way. In fact. if (5.9) holds for some C^1-path $x : J - E_x$. then it follows from the integral mean value theorem that $F_1(x(t)) = F_1(x(t_0))$ for any fixed $t_0 \in J$. Hence $t \in J \quad - (x(t). x'(t))$ is a path on the tangent bundle of the manifold M_b specified by $b = F_1(x_0)$. This suggests that we imbedd (5.7) into the family of DAEs

$$(5.10) \qquad\qquad \begin{aligned} F_1(x) &= b, \ b \in F_1(E_x) \\ F_2(x. x'. z) &= 0 \end{aligned}$$

indexed by the vectors of $F_1(E_x)$.

Let $x : J - E_x$, $z : J - E_z$ be a C^1-solution of a member of (5.10). Then for any point $(x_0, p_0, z_0) = (x(t_0). x'(t_0). z(t_0))$, $t_0 \in J$ the value $b = F_1(x_0)$ uniquely specifies the particular DAE. But, in addition, (x_0, p_0, z_0) must satisfy $DF_1(x_0)p_0 = 0$, as well as $F_2(x_0, p_0. z_0) = 0$. This suggests the definition of the C^1-map

$$(5.11) \qquad H : E_2 - \mathrm{R}^s \times \mathrm{R}^r. H(x, p, z) = \begin{pmatrix} DF_1(x)p \\ F_2(x, p, z) \end{pmatrix}, \ \forall (x, p, z) \in E_2$$

as the initial data map of the family (5.10).

For any given $(x, p, z) \in E_2$ the derivative $D_{p,z}H$ of H with respect to $(p. z)$ is exactly the linear operator in condition (B3). The nonsingularity of $D_{p,z}H$ implies the following result:

Lemma 5.1 *For any $(x_0, p_0, z_0) \in K$ we have an open neighborhood $U = S_x \times S_p \times S_z$ in E_2, and unique C^1-maps $\eta : S_x \to S_p$, $\zeta : S_x \to S_z$, with $\eta(x_0) = p_0$, $\zeta(x_0) = z_0$. such that for any $x \in S_x$ the only solution $(p, z) \in S_p \times S_z$ of $H(x, p, z) = 0$ is $p = \eta(x)$, $z = \zeta(x)$.*

Thus, on the open neighborhood $S_x \subset E_x$ of x_0,

$$(5.12) \qquad\qquad \pi : S_x \to TS_x, \ \pi(x) = (x, \eta(x)), \ x \in S_x.$$

constitutes a C^1-vectorfield. Since $D_{p,z}H \in \mathrm{Isom}(\mathrm{R}^{n+m})$, the mapping H is a submersion and hence the solution set

$$(5.13) \qquad\qquad K = \{(x, p, z) \in E_2; \ H(x, p, z) = 0\}$$

is an n-dimensional C^1-submanifold of R^{2n+m}. This manifold turns out to be state space of the family of DAEs (5.10). In fact, it can be shown that π is tangential to the constraint manifold through x; that is,

$$\pi(x) \in T_x M_b, \ b = F_1(x) \ \forall x \in S_x,$$

and that for any solution $x : J \longrightarrow S_x$ of the explicit ODE

$$(5.14) \qquad\qquad\qquad\qquad x' = \eta(x),$$

we obtain the C^1-solution $t \in J \longrightarrow (x(t), \zeta(x(t)))$ of the member of (5.10) specified by $b = F_1(x(t_0))$ for arbitrary fixed $t_0 \in J$.

Thus, from the standard existence theory of initial value problems for ODEs we obtain the following result:

Theorem 5.2 *Suppose that the conditions (B1,2,3) hold and that K is non-empty. Then any point $(x_0, p_0, z_0) \in K$ has an open neighborhood $U \equiv S_x \times S_p \times S_z \subset E_2$ such that for any $x_c \in S_x$ there is exactly one point $(x_c, p_c, z_c) \in K \cap U$. Moreover, for any $x_c \in S_x$ there exists a unique, maximally extended C^1-solution $x : J \longrightarrow S_x$, $z : J \longrightarrow S_z$, on some open interval J with $0 \in J$, of the DAE (5.10) specified by $b = F_1(x_c)$ which satisfies $x'(J) \subset S_p$ and the initial conditions $x(0) = x_c$, $x'(0) = p_c$, $z(0) = z_c$.*

We refer to [41] for a globalization of this result based on the techniques from the theory of covering spaces mentioned at the end of Section 4.

The result extends to the second order DAEs (4.3). In analogy with (B1,2,3) we suppose that the problem

$$(5.15) \qquad\qquad F(x, x', x'', z) = \begin{pmatrix} F_1(x) \\ F_2(x, x', x'', z) \end{pmatrix} = 0$$

satisfies the conditions:

(C1) $F_1 : E_x \in \mathbf{R}^n \longrightarrow \mathbf{R}^s$ is C^3,

(C2) $F_2 : E_2 = E_x \times E_y \times E_q \times E_z \in \mathbf{R}^{3n+m} \longrightarrow \mathbf{R}^r$ is C^1,

(C3) $\begin{pmatrix} DF_1(x) & 0 \\ D_p F_2(x, y, q, z) & D_z F_2(x, y, q, z) \end{pmatrix} \in \mathrm{Isom}(\mathbf{R}^{m+n})$ for each $(x, p, q, z) \in E_2$,

where $s \le n \le s + r = n + m$ and E_x, E_y $E_q \subset \mathbf{R}^n$, $E_z \subset \mathbf{R}^m$ again are non-empty, open sets.

It is natural to reduce (5.15) to a first order system by introducing a new variable y and adding the equation $x' = y$. Then it turns out that, with (x, y) as new differential variable, the resulting system constitutes a DAE of the form (5.7) for which (B1) and (B2) are valid. However, in general, (B3) does not hold which is hardly surprising since we should expect (5.15) to induce local second order vectorfields instead of the local first order fields (5.12).

If $x : J \to E_x$ is a C^2-path on M_b for some $b \in F_1(E_x)$: that is. if $F_1(x(t)) = b$ for $t \in J$. then for all $t \in J$ the we must have

$$DF_1(x(t))x'(t) = 0,$$

as well as

$$(5.16) \qquad D_2F_1(x(t))x''(t) + DF_1(x(t))(x'(t), x'(t)) = 0.$$

This shows that $t \in J \to ((x(t), x'(t)), (x'(t), x''(t)))$ is a path on the second tangent bundle $T^2 M_b$ of M_b. Conversely, by the integral mean-value theorem we obtain the following result:

Lemma 5.2 Let $x : J \to E_x$ be any C^2-path that satisfies (5.16). If there exists a t_0 in J such that $DF_1(x(t_0))x'(t_0) = 0$ and therefore $(x(t_0), x'(t_0)) \in TM_b$ for $b = F_1(x(t_0))$. then $((x(t), x'(t)), (x'(t), x''(t))) \in T^2 M_b$, for all $t \in J$.

As in the first order case this suggests that we imbedd (5.15) into the family of DAEs

$$(5.17) \qquad \begin{aligned} F_1(x) &= b, \quad b \in F_1(E_x) \\ F_2(x, x', x'', z) &= 0 \end{aligned}$$

and that we define the C^1-initial-data map:

$$(5.18) \qquad H : E_2 \to \mathbf{R}^s \times \mathbf{R}^r$$

$$H(x, y, q, z) = \begin{pmatrix} DF_1(x)q + D_2F_1(x)(y, y) \\ F_2(x, y, q, z) \end{pmatrix}, \forall (x, y, q, z) \in E_2.$$

By (C3) we have $D_{q,z}H(x, y, q, z) \in \mathrm{Isom}(\mathbf{R}^{s+r}, \mathbf{R}^{n+m})$ for $(x, y, q, z) \in E_2$, and hence H is a submersion and the solution set

$$(5.19) \qquad K = \{(x, y, q, z) \in E_2; \quad H(x, y, q, z) = 0\}$$

is a $2n$-dimensional C^1-submanifold of $\mathbf{R}^{3n+m}$ and the following result holds:

Lemma 5.3 For any $(x_0, y_0, q_0, z_0) \in K$ there exists an open neighborhood $U = S_x \times S_y \times S_q \times S_z$ in E_2, and unique C^1-maps $\eta : S_0 \equiv S_x \times S_y \to S_p$, $\zeta : S_x \to S_z$, with $\eta(x_0, y_0) = q_0$, $\zeta(x_0, y_0) = z_0$, such that for any given $(x, y) \in S_0$ the only solution $(q, z) \in S_p \times S_z$ of $H(x, y, q, z) = 0$ is given by $q = \eta(x, y)$, $z = \zeta(x, y)$.

Using this lemma we can now define on the open neighborhood S_0 of (x_0, y_0) the second-order C^1-vectorfield

$$(5.20) \qquad \pi : U \in TS_x \to T^2 S_x, \quad \pi(x, y) = ((x, y), (y, \eta(x, y))), \quad \forall (x, y) \in S_0.$$

Then. for any $(x,y) \in S_0$ such that $DF_1(x)y = 0$, it follows that $\pi(x,y) \in T^2 M_b$ for $b = F_1(x)$. Moreover, if $(x,y) : J \to S_0$ is any solution of the explicit ODE-system

$$(5.21) \qquad\qquad x' = y, \quad y' = \eta(x,y),$$

satisfying $DF_1(x(t_0))y(t_0) = 0$ for some $t_0 \in J$, then $t \to (x(t), \zeta(x(t)))$ is a C^1-solution of (5.17) for $b = F_1(x(t_0))$.

Thus the standard solution theory provides here the following local existence result:

Theorem 5.3 *Suppose that the conditions (C1.2.3) hold and that K is non-empty. Then any $(x_0, y_0, p_0, z_0) \in K$ has an open neighborhood $U = \cdots \times S_y \cdot S_q \times S_z$ in E_2 such that for any $(x_c, y_c) \in S_0 = S_x \times S_y$ there is exactly one po: $(x_c, y_c, p_c, z_c) \in K \cap U$. Moreover, for any $(x_c, y_c) \in S_0$ with $DF_1(x_c)y_c = 0$ there exists a unique, maximally extended C^1 solution $x : J \in S_x$, $z : J \in S_z$ of (5.17) for $b = F_1(x_c)$ on some open interval $J \subset R^1$ containing the origin which satisfies $x'(J) \subset S_y$, $x''(J) \subset S_q$ and the initial condition $x(0) = x_c$, $x'(0) = y_c$, $x''(0) = q_c$, $z(0) = z_c$.*

Once again covering-space theory can be used to globalize this result (see [41]).

6 Numerical Methods for DAEs

6.1 Application of ODE Methods

The most frequently used approach to the computational solution of DAEs is the application of standard ODE methods. This idea appears to be due to W.Gear [17] who proposed the use of backward-difference (BDF) methods in the form developed for stiff ODEs.

Briefly, in an m-step BDF-method the derivative x' of the unknown function at the time t_k, $k \geq m$, is approximated by the derivative of the interpolation-polynomial through (x_k, t_k) and m earlier computed points (x_{k-i}, t_{k-i}), $i = 1, 2, ..., m$. Hence, in the case of the implicit initial value problem

$$(6.1) \qquad\qquad F(t, x, x') = 0, \quad x(t_0) = x_0, \quad x'(t_0) = p_0$$

the determination of x_k requires the solution of the nonlinear system of equations

$$(6.2) \qquad\qquad F\left(t_k, x_k, \frac{1}{h_k} \sum_{i=0}^{m} \alpha_{k,i} x_{k-i}\right) = 0.$$

Here $\alpha_{k,i}$, $i = 0, 1, \dots, m$ are the coefficients of the BDF formula at the k-th step which, of course, depend on k unless the stepsizes $h_i = t_i - t_{i-1}$ remain constant. For $m < 7$

the m-step BDF methods are known to be stable when applied to ODEs and hence we assume from now on that $1 < m \leq 6$.

When (6.1) represents a DAE; that is, when the constant-rank condition (3.2) holds. then the validity and performance of the process depends on several factors. In particular. the initial value problem (6.1) has to possess a solution and for $m > 1$ the required m additional starting points have to approximate this solution. Moreover. at each step the nonlinear system (6.2) has to have a feasible solution which is computable by a suitable iterative process such as some form of Newton's method. In general. the answers to these questions depend strongly on the index of the DAE.

For simplicity. we restrict ourselves here to the semi-implicit equation (4.5); that is. to the initial value problem

$$(6.3) \qquad F(x. x') = \begin{pmatrix} F_1(x) \\ F_2(x, x') \end{pmatrix}, \quad x(0) = x_0. \ x'(0) = p_0.$$

If the conditions (A1,2,3) hold. then. for (x_0, p_0) in the set E_A deined by (4.15). Theorem 4.1 ensures that (6.3) has a unique (local) solution. Moreover. on E_A the derivative $D_p \hat{F}$ of the reduced mapping (4.10) is non-singular.

The nonlinear system. to be solved at each step of the process. can be written in the form

$$G(x) = \begin{pmatrix} F_1(x) \\ hF_2(x, \frac{1}{h}(x - w)) \end{pmatrix} = 0,$$

where w incorporates all information at the earlier computed points. All basic forms of Newton's method are locally convergent if the derivative DG is non-singular at the desired solution. Hence we are interested in the non-singularity of the matrix

$$\begin{pmatrix} DF_1(x) \\ D_pF_2(x,p) + hD_xF_2(x,p) \end{pmatrix} = D_p\hat{F}(x,p) + h \begin{pmatrix} 0 \\ D_xF_2(x.p) \end{pmatrix}.$$

which, by definition of E_A, is clearly guaranteed for all (x, p) in some open neighborhood of any point on the exact solution in E_A and for all sufficiently small h.

Under these conditons it can be shown that when an m-step BDF method is used for the computational solution of (6.3), together with a fixed and sufficiently small stepsize h, then the convergence of the approximate points to the exact solution is of order $O(h^m)$ provided that all initial points are correct to order $O(h^m)$ and stopping criteria of order $O(h^{m+1})$ are applied in the Newton process at each step. A proof of this result for the general system (6.1) may be found in [3] where also its extension to the case of variable steps is discussed. These results about BDF methods for index-one systems form the theoretical basis for several highly successful numerical DAE-solvers notably the widely used codes DASSL [30] and LSODI [25].

Besides BDF-methods also other multistep have been considered in the DAE literature. In particular. an extensive analysis of general linear multistep methods for the index-one case is given in [22].

The situation changes considerably when the DAE has index higher than one and two basic difficulties arise. The first derives from the fact that for any particular multistep method [2] there exist DAEs with index exceeding one for which the method is unstable. [21]. The second difficulty is the appearance of a transient deterioration of the discretization errors following any change of the step-size in the method. This type of "boundary layer" was observed by several authors. see. e.g., [29,43].

More specifically. in [43] it was proved that when an m-step. fixed-stepsize BDF method is applied to a linear DAE (4.1) with index $\kappa \geq 1$. then the process converges with order $O(h^m)$ after $(\kappa - 1)m - 1$ steps. Moreover. in [19] it was shown that when variable stepsizes are used and the ratios of adjacent steps remain bounded then the global error has order $O(h_{max}^\mu)$ where $\mu = \min(m, m - \kappa - 2)$. Hence. for instance. for an index-three system the use of the implicit Euler method with variable stepsizes may lead to errors of order $O(1)$. However. note that for index-two problems we have $\mu = m$ and. in fact. it turns out. [2.4.27], that for semi-implicit systems (6.3) of index two after $m - 1$-steps the m-step BDF method with fixed steps is globally convergent of order $O(h^m)$ provided again that the initial points are correct to order $O(h^m)$ and the stopping-criteria of the iterative process at each step have order $O(h^{m-1})$. But. nevertheless. these iteretive methods may well converge very poorly in the beginning steps. The variable-stepsize case of this result is discussed in [20].

Besides multistep methods various one-step methods have also been considered for the computational solution of DAEs. In particular, there exists a large literature on the use of implicit Runge-Kutta (IRK) methods. Any such method can be characterized by its Butcher tableau

$$
\begin{array}{c|cccc}
c_1 & a_{11} & a_{12} & \cdots & a_{1m} \\
c_2 & a_{21} & a_{22} & \cdots & a_{2m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
c_m & a_{m1} & a_{m2} & \cdots & a_{mm} \\
\hline
 & b_1 & b_2 & \cdots & b_m
\end{array}
$$

see e.g. [7]. When applied to the DAE (6.1) the basic algorithm assumes the form

$$(i) \text{ solve} \qquad F(t_{k-1} + c_i h, x_{k-1} + h \sum_{j=1}^{m} a_{ij} Y_i, Y_j) = 0, \quad i = 1, 2, \ldots, m$$

$$\text{for } Y_1, \ldots, Y_m \in \mathbb{R}^n$$

---

[2] In fact this holds also for Runge-Kutta methods.

$$(ii) \quad \text{set} \qquad x_k = x_{k-1} - h \sum_{j=1}^{m} b_j Y_j.$$

Implicit Runge-Kutta methods are useful for generating accurate initial data for higher order multistep methods; they are also advantageous for problems with multiple discontinuities. In general, the nonlinear system arising at each step has dimension mn and may be very costly to solve unless A has special properties. Thus the complexity of theses processes depends strongly on the form of the coefficient matrix $A = (a_{ij})$. Some important special cases include the DIRK-methods for which A is block-lower-triangular with equal diagonal blocks as well as the SIRK-methods where A has one real eigenvalue.

In the numerical integration of stiff ODEs it has become well-known that the computed solution often exhibits a disappointingly low accuracy when compared with the order of consistency of the method. For Runge-Kutta methods applied to a class of stiff linear ODEs this was first observed in [34] where it was noted that for stiff problems the order of consistency should not be based on the classical Lipschitz condition. Instead in [15] and several subsequent papers (see also the monograph [12]) one-sided Lipschitz conditions were used to introduce the concepts of B-consistency and B-convergence which provide order results that correspond more closely to the observed behavior for stiff ODEs.

For an IRK method let the stage order be the largest integer $r \geq 1$ such that the conditions

$$\sum_{j=1}^{m} a_{ij} c_j^{k-1} = \frac{1}{k} c_i^k, \quad i = 1, 2, ..., m,$$

are valid for $k = 1, 2, \ldots, r$. Moreover, define the quadrature order as the largest integer $q \geq 1$ for which the conditions

$$\sum_{j=1}^{m} b_j c_j^{k-1} = \frac{1}{k},$$

hold for $k = 1, 2, \ldots, q$. Then $\bar{p} = \min(r, q)$ is called the internal stage order and for $q > \bar{p}$ the classical nonstiff ODE order p satisfies $q \geq p \geq \bar{p} + 1$. There are examples of stiff ODEs and IRK-methods where $p > \bar{p}$ and the observed order of convergence equals the internal stage order $\bar{p}$ (see e.g. [12]).

This behavior is mirrored in the application of IRK methods to DAEs. In fact, DAEs have a close relationship with stiff ODEs, as is suggested, for instance, by the fact that singularly perturbed systems

$$x' = -1(x, y), \quad \epsilon y' = f_2(x, y)$$

with small $\epsilon > 0$ become a DAE for $\epsilon = 0$.

Consider again (6.3) subject to the conditions (A1.2.3) and with $(x_0, p_0) \in E_A$. For the approximation of the solution in E_A we consider an IRK method with a non-singular coefficient matrix A such that

$$1 - b^T A^{-1} e_i < 1, \quad e = (1, 1, \ldots, 1)^T \in \mathbb{R}^m,$$

which means that the method is A-stable (see e.g. [7]). If $q > \tilde{p} > 1$ then for a constant (sufficiently small) step-size $h > 0$ the global error is at least of order $O(h^{\tilde{p}-1})$ provided any error in the initial point and the termination criterion for Newton's method are of the same order.

A proof of this result is given in [3] where also exa⋯s of problems are found for which the achieved orders are higher than the stated ⋯r bound. In essence, the results also carry over to semi-implicit index-two problem⋯ see again [3] and also [5]).

Various other one-step methods have been used for the computational solution of DAEs. This includes, for example, the Runge-Kutta-Rosenbrock methods considered in [42] and the extrapolation methods studied by Deuflhard et al (see e.g. [13] or [14]). We shall not enter here into any further detail.

6.2 Local Parametrizations

In this section we turn to a local parametrization approach suggested by the existence results of the earlier sections. It was introduced in [41] and considered further in [32.31], and is related to the generalized coordinate partitioning technique used in the numerical solution of Euler-Lagrange equations by E. Haug et al (see [28,48])

Consider first the system (4.2) subject to the conditions (B1,2,3), and, more specifically, suppose that we are in the setting of Theorem 5.3. Then, for given $(x_0, p_0, z_0) \in K$ we wish to compute the C^1 solution $x : J \to S_x$, $z : J \to S_z$, of the DAE (5.17) specified by $b = F_1(x_0)$ that satisfies $x'(J) \subset S_p$ as well as the initial conditions $x(0) = x_0$, $x'(0) = p_0$, $z(0) = z_0$. For any $x \in S_x$ the unique solution p.z of the equations

$$DF_1(x)p = 0, \quad F_2(x, p, z) = 0, \quad (x, p, z) \in K$$

is provided by the values $p = \eta(x)$ and $z = \zeta(x)$ of the mappings of Lemma 5.1. As we saw, once a procedure is available for computing $\eta(x)$ and $\zeta(x)$ for any needed $x \in S_x$, the problem of solving (5.17) in S_x reduces to that of solving the explicit ODE

$$(6.4) \qquad\qquad x = \eta(x), \quad x \in S_x.$$

Since the desired solution $x : J \to S_x$ of (6.4) through x_0 has to remain on the constraint manifold M_b through x_0 it is natural to work with a local coordinate system on M_b. For this we use a simple class of such coordinate systems applied earlier in other differential-geometric numerical methods (see e.g. [39,40]). Let $x_c \in M_b$ be any

point on the $n - s$-dimensional constraint manifold M_b through x_0 and consider any linear subspace $T \subset R^n$, with $\dim T = n - r$, such that

$$(6.5) \qquad\qquad T^- \cap \ker DF(x_c) = \{0\}.$$

If $A \in L(R^{n-s}, R^n)$ is any matrix with orthonormal columns which span T, then (6.5) is equivalent with the assumption that

$$(6.6) \qquad\qquad \begin{pmatrix} DF_1(x_c) \\ A^T \end{pmatrix} \in \mathrm{Isom}(R^n).$$

Hence there exist an open neighborhood V_0 of the origin in R^{n-s} such that for all $u \in V_0$ the system

$$\begin{pmatrix} F_1(x) \\ A^T(x - x_c) \end{pmatrix} = \begin{pmatrix} 0 \\ u \end{pmatrix}$$

has a unique solution $x = \Psi(u) \in R^n$. It is readily seen that the resulting mapping Ψ is a C^1 diffeomorphism from V_0 onto the relatively open neighborhood $\Psi V_0 \subset M_b$ of x_c. This is the desired local coordinate mapping. Note that with $\omega(u) = \Psi(u) - x_c - Au$ we have $A^T \omega(u) = 0$, and thus

$$\Psi : V_0 \to R^n, \quad \Psi(u) = x_c + Au + \omega(u) \in M_b, \quad u \in V_0.$$

shows that the point $x = \Psi(u)$ on M_b is obtained by adding to x_c the vector $Au \in T$ and the orthogonal correction $\omega(u) \in T^\perp$. The local coordinate system at the point x_c of M_b is completely determined by the matrix A and hence we shall also speak of the local coordinate system induced by that matrix.

In practice, it is often useful to work with local coordinate mappings defined by the tangent space $T = \ker DF(x_c)$ at $x_c \in M_b$ (see e.g. [40]). Suppose that we compute the QR-factorization

$$DF_1(x_c)^T = (Q_1, Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix}$$

where the matrices $Q_1 \in L(R^s, R^n)$ and $Q_2 \in L(R^{n-s}, R^n)$ have orthonormal columns and $R \in L(R^s, R^s)$ is upper-triangular and nonsingular. Then Q_2 can be used as the basis matrix A of T while the columns of Q_1 span $T^\perp$.

A second, practically useful choice of a local-coordinate space T and its basis matrix A consists in determining a permutation $e^{j_1}, e^{j_2}, \ldots, e^{j_n}$ of the standard basis of R^n such that the matrix

$$A = (e^{j_{s+1}}, e^{j_{s+2}}, \ldots, e^{j_n})$$

satisfies (6.6). This choice partitions the components of the vector $x = (x_1, x_2, \ldots, x_n)$ into a vector $(x_{j_{s+1}}, x_{j_{s+2}}, \ldots, x_{j_n})$ of independent coordinates and the complementary

vector $(x_{j_1}, x_{j_2}, \ldots, x_{j_s})$ of dependent coordinates. This is the choice underlying the mentioned generalized coordinate partitioning approach of E.Haug et al (loc. cit.).

Once a local coordinate system has been chosen then it can be shown (see [41]) that the ODE (6.4) has the local representation

$$(6.7) \qquad\qquad u' = A^T \eta(\Psi(u)), \quad u \in V_0 \subset R^{n-s}.$$

This is an $(n-s)$-dimensional explicit ODE without constraints to which any standard ODE solver can be applied as long as the computed points remain in V_0.

This local coordinate approach can also be carried over to the second order DAEs (4.3). As before, suppose analogously that we are in the setting of Theorem 5.3. Let $(x_0, y_0, p_0, z_0) \in K$ be a given point for which $DF_1(x_0)y_0 = 0$ and suppose that we wish to compute the C^1-solution $x : J \to S_x$, $z : J \to S_z$ of (5.17) specified by $b = F_1(x_0)$ that satisfies $x'(J) \subset S_y$, $x''(J) \subset S_q$ and the initial conditions $x(0) = x_0$, $x'(0) = y_0$. $x''(0) = q_0$, $z(0) = z_0$. For any $(x, y) \in S_0 \equiv S_x \times S_y$ the unique solution q, z of

$$DF_1(x)q + D_2F_1(x)(y, y) = 0, \quad F_2(x, y, q, z) = 0, \quad (x, y, q, z) \in K$$

is given by the values $q = \eta(x, y)$ and $z = \zeta(x, y)$ of the mappings of Lemma 5.3. Thus, when a method for evaluating $\eta(x, y)$ and $\zeta(x, y)$ is available, then the problem of solving (5.17) in S_0 is reduced to that of solving the explicit first order system

$$(6.8) \qquad\qquad x' = y, \quad y' = \eta(x, y).$$

As we know the desired solution $(x, y) : J \to S_0$ of (6.10) through (x_0, y_0) remains on the tangent bundle TM_b of the constraint manifold M_b through x_0. Thus we have to work here with a local coordinate system on TM_b. As before, let the matrix $A \in L(R^{n-s}, R^n)$ induce a local coordinate system at the point $x_c \in M_b$ of M_b. Then

$$(6.9) \qquad \Theta : V_0 \times R^{n-s} \to TM_b, \quad \Theta(u, v) = (\Psi(u), D\Psi(u)v), \quad u \in V_0, \ v \in R^{n-s}$$

defines a local coordinate system on TM_b. By restricting V_0, if necessary, we can choose some neighborhood U_0 of the origin of $V_0 \times R^{n-s}$ which Θ maps into E_0.

In this local coordinate system the differential equations (6.8) assume the local form

$$(6.10) \qquad\qquad u' = v, \quad v' = A^T \eta(\Psi(u), D\Psi(u)v), \quad (u, v) \in U_0$$

and hence, once again a standard ODE solver can be applied as long as the computed points remain in U_0

6.3   Euler-Lagrange Equations

In this section we sketch briefly how a multistp method might be applied when the local parametrization approach is used in the numerical solution of the Euler-Lagrange equations (2.3), (2.4); that is the constrained equations of motion

$$(6.11) \qquad \Phi(x,t) = 0$$
$$M(x,t)x'' - D_x\Phi(x,t)^T z = Q(x,x',t).$$

For further detail we refer to [24,33].

In the case of (6.11) it is easily seen that (C3) is equivalent with the assumption

$$(6.12) \qquad \mathrm{rank}\, D_x\Phi(x,t) = s, \quad y^T M(x,t)y > 0, \ \forall y \in \ker D_x\Phi(x,t)$$

which is equivalent with the non-singularity of the matrix of the linear system

$$(6.13) \qquad \begin{pmatrix} M(x,t) & D_x\Phi(x,t)^T \\ D_x\Phi(x,t) & 0 \end{pmatrix} \begin{pmatrix} q \\ z \end{pmatrix} = \begin{pmatrix} Q(x,y,t) \\ g(x,y,t) \end{pmatrix}.$$

Thus under the condition (6.12) the general existence theory applies to (6.11), (see e.g. [41]).

For given (x,y,t), set

$$g(x,y,t) = -(D_{xx}^2\Phi(x,t)(y,y) - D_{xt}^2\Phi(x,t)y - D_{tt}^2\Phi(x,t)).$$

in (6.13) and let $q = \eta(x,y,t)$, $z = \zeta(x,y,t)$ be the unique solution of that system. Then the problem of solving (6.11) is reduced to solving the explicit first order system

$$(6.14) \qquad x' = y, \quad y' = \eta(x,y,t).$$

As in the previous section let $A \in L(\mathrm{R}^{n-s}, \mathrm{R}^n)$ induce a local coordinate system at the current point of M_b and introduce the corresponding local coordinate system (6.9) on the tangent bundle TM_b. Then it follows readily that the local representation of (6.14) is given by

$$(6.15) \qquad u' = v, \quad v' = A^T\eta(\Psi(u,t), D_u\Psi(u,t)v + D_t\Psi(u,t), t).$$

Suppose that for its solution we use a (consistent) explicit multistep method of the form

$$u_k = \sum_{j=1}^m \alpha_j u_{k-j} + h\sum_{j=1}^m \beta_j v_{k-j}, \quad v_k = \sum_{j=1}^m \alpha_j v_{k-j} - h\sum_{j=1}^m \beta_j v'_{k-j}$$

with constant step $h > 0$. For the computation it is advantageous not to work with the local variables u, v but to transform all formulas immediately back to the original variables x, y.

If the approximations $x_{k-j}, y_{k-j}, y'_{k-j}, z_{k-j}$, $j = 1, 2, \ldots, m$ of the solution are already available, then the algorithm for computing x_k, y_k, y'_k, z_k has the form:

(i) Set $t_k = t_{k-1} + h$;

(ii) Evaluate

$$a_k = A^T \{ \sum_{j=1}^{m} \alpha_j x_{k-j} - h \sum_{j=1}^{m} \beta_j y_{k-j} \} \quad a_k' = A^T \{ \sum_{j=1}^{m} \alpha_j y_{k-j} - h \sum_{j=1}^{m} \beta_j y_{k-j}' \} :$$

(iii) Solve the nonlinear system

$$\begin{pmatrix} \Phi(x, t_k) \\ A^T x \end{pmatrix} = \begin{pmatrix} 0 \\ a_k \end{pmatrix}$$

and set $x_k = x$;

(iv) Solve the linear system

$$\begin{pmatrix} D_x \Phi(x_k, t_k) \\ A^T \end{pmatrix} y = \begin{pmatrix} -D_t \Phi(x_k, t_k) \\ a_k' \end{pmatrix}$$

and set $y_k = y$;

(v) Solve the linear system

$$\begin{pmatrix} M(x_k, t_k) & D_x \Phi(x_k, t)^T \\ D_x \Phi(x, t) & 0 \end{pmatrix} \begin{pmatrix} w \\ z \end{pmatrix} = \begin{pmatrix} Q(x_k, y_k, t_k) \\ g(x_k, y_k, t_k) \end{pmatrix}$$

and set $y_k' = w$ and $z_k = z$.

In stage (ii) the multistep formula is evaluated in terms of the original variables x, y. Then the stages (iii) and (iv) determine the local coordinate mapping (6.15) and finally in stage (v) the linear system (6.13) is solved to obtain the accelerations y_k' and the algebraic variable z_k defining the constraint force.

In stage (iii) a chord-Newton process can be used involving the matrix obtained in stage (iv) of the previous solution-step. When the computed points leave the domain of validity of the current local coordinate system, then the matrix A has to be updated. The need for this can be detected by monitoring the number of iteration-steps of the nonlinear solver in stage (ii) or the condition of the linear system in stage (iii).

When an implicit multistep method is to be used then all three steps (iii)-(v) have to be combined into one. Now special attention has to be given to the inherent struc· :re of the resulting large nonlinear system in order to keep the computational complexity at an acceptable level. For some detail we refer to [31] where also a numerical example is given.

7 Appendix

In this Appendix we collect some background material used throughout the presentation. For further details, especially on the differential geometric aspects, we refer to standard text such as [44] or [1].

As usual, a mapping $F : U \to \mathrm{R}^m$ on the open set $U \subset \mathrm{R}^n$ is of class C^r, $r \geq 0$, on U if F is continuous and for $r > 0$ all its partial derivatives up to and including order r exist and are continuous on U. More generally, a map $F : S \to \mathrm{R}^m$ on an arbitrary set $S \subset \mathrm{R}^n$ is of class C^r if for each $x \in S$ there exists an open set $U \subset \mathrm{R}^n$ containing x and a C^r-mapping $\hat{F} : U \to \mathrm{R}^m$ that coincides with F throughout $U \cap S$. A map $F : S \subset \mathrm{R}^n \to T \subset \mathrm{R}^m$ is a homeomorphism between the sets S and T if F is a one-to-one mapping from S onto T and both F and its inverse $F^{-1} : T \to S$ are continuous. Finally, a map $F : S \subset \mathrm{R}^n \to T \subset \mathrm{R}^m$ is a C^r-diffeomorphism if F is a homeomorphism between S and T and if both F and F^{-1} are of class C^r.

A subset $M \subset \mathrm{R}^n$ is a d-dimensional C^r-sub-manifold of R^n if for each point $x \in M$ there exists an open set $U \subset \mathrm{R}^n$ containing x such that the neighborhood $U \cap M$ of x on M is C^r-diffeomorphic to an open subset V of R^d. Any particular such diffeomorphism $\varphi : U \cap M \to V$ is called a chart and its inverse a local coordinate system on $U \cap M$.

By this definition any open subset $U \subset \mathrm{R}^n$ is an n-dimensional C^∞-sub-manifold of R^n. The tangent space $T_x U$ of this manifold U at any point $x \in U$ is defined as the n-dimensional linear space $\{x\} \times \mathrm{R}^n$, and its tangent bundle TU is the $2n$-dimensional submanifold $U \times \mathrm{R}^n$ of R^{2n}.

Let $F : U \subset \mathrm{R}^n \times \mathrm{R}^m$, $n > m$, be some C^r-mapping, $r \geq 1$, on the open set $U \subset \mathrm{R}^n$. A point $x \in U$ is a regular point of F if $\dim DF(x)\mathrm{R}^n = m$; that is, if the derivative $DF(x)$ has full rank m. If all points of a set $S \subset \mathrm{R}^n$ are regular points then F is a submersion on S. A point $b \in \mathrm{R}^m$ is a regular value of F if all points of the inverse image $F^{-1}(b) = \{x \in U, F(x) = b\}$ are regular; that is, if F is a submersion on $F^{-1}(b)$. A fundamental result then states that for any regular value $b \in \mathrm{R}^m$ the inverse image $M_b = F^{-1}(b)$ is either empty or a $p = (n - m)$-dimensional C^r-sub-manifold of R^n. The tangent space $T_x M_b$ at any point x of this manifold M_b may be identified with the set

$$T_x M_b = \{(x,p) \in T_x \mathrm{R}^n; DF(x)p = 0\}.$$

Clearly, $T_x M_b$ is a p-dimensional linear subspace of the n-dimensional linear space $T_x \mathrm{R}^n$. The tangent bundle TM_b of M_b is the disjoint union of all tangent spaces $T_x M_b$ for $x \in M_b$; that is,

$$TM_b = \{(x,p) \in TU; F(x) = b, DF(x)p = 0\},$$

and TM_b is a $2d$-dimensional C^{r-1}-sub-manifold of $T\mathrm{R}^n$. Evidently, then the tangent bundle of TM_b is the $4d$-dimensional C^{r-2}-sub-manifold

$$T^2 M_b = \{((x,y)(p,q)) \in T^2 U; F(x) = b, DF(x)p = 0, DF(x)q + D2F(x)(y,p) = 0\}.$$

of R^{4n}.

A C^s-vectorfield on some open subset $U \subset R^n$ is a C^s-mapping on U such that

(7.1) $$\pi : U \to TU; \quad \pi(x) = (x, \eta(x)), \quad \forall x \in U.$$

An integral curve of π through a point $x_0 \in U$ is any C^s-path $x : J \to U$, defined on an open interval $J \subset R^1$ containing the origin, for which $x(0) = x_0$ and $(x(t), x'(t)) = \pi(x(t))$ for $t \in J$; that is, which solves the initial value problem

$$x' = \eta(x), \quad x \in U, \quad x(0) = x_0.$$

For a vectorfield (7.1) of class C^s, $s \geq 1$, on the (non-empty) open subset $U \subset R^n$, the following results hold:

(i) There exists a C^s-integral curve $x : J \to U$ of π through each $x \in U$ defined on an open interval J. Moreover, any two such curves are equal on the intersection of their domains.

(ii) The union of the domains of all integral curves of π through a point $x \in U$ is an open, possibly unbounded interval J_x^*. There exists a C^s-integral curve $x^* : J_x^* \to U$, of π through x and J_x^* is the largest interval on which such an intergal curve exists.

(iii) The set $D(\pi) = \{(t, x) \in R^1 \times U; \ t \in J_x^*\}$ is open in $R^1 \times U$ and contains $\{0\} \times U$. Moreover, the global flow $\xi : D(\pi) \to U$, $\xi(t, x) = x^*$, $t \in J_x^*$ of π is of class C^s on $D(\pi)$.

Consider now a second order ODE $x'' = \eta(x, x')$, $x \in U$ where η is of class C^s on some open set $E \subset R^{2n}$. When this problem is written in the first order form $x' = y$, $y' = \eta(x, y)$, $(x, y) \in E$, then we encounter a vector-field of the form

(7.2) $$\pi : E \subset TU \to T^2U; \quad \pi(x, y) = ((x, y), (y, \eta(x, y))), \quad \forall(x, y) \in E.$$

Note that the second and third component of the image vector are identical; in other words, (7.2) represents a sub-class of all tangential vector fields on TU, namely, the vector fields that are consistent with the second order ODE.

An integral curve of (7.2) through a point $(x_0, y_0) \in E$ is now a C^s-path $x : J \to R^n$, defined on some open interval $J \subset R^1$ containing the origin, for which $(x(t), x'(t)) \in E$ and $((x(t), x'(t)), (x'(t), x''(t))) = \pi((x(t), x'(t)))$ for all $t \in J$; that is, which is a solution of the original second order ODE.

References

[1] R. ABRAHAM. J. E. MARSDEN. AND T. RATIU. *Manifolds. Tensor Analysis. and Applications.* Springer Verlag. New York. NY. 1988.

[2] K. E. BRENAN. *Stability and Convergence of Difference Approximations for Higher-Index Differential-Algebraic Systems with Applications in Trajectory Control.* PhD thesis, University of California at Los Angeles, 1983.

[3] K. E. BRENAN, S. L. CAMPBELL. AND L. R. PETZOLD. *Numerical Solution of Initial Value Problems in Differential-Algebraic Equations.* North-Holland Publ. Co., New York, NY. 1989.

[4] K. E. BRENAN AND B. E. ENGQUIST. *Backward differentiation approximations of nonlinear differential/algebraic equations.* Math. of Comp.. 51 (1988), pp. 659–676 and S7–S16.

[5] K. BURRAGE AND L. R. PETZOLD, *On order reduction for Runge-Kutta methods applied to differential-algebraic systems and to stiff systems of ODEs.* Tech. Report UCRL-98046, Lawrence Livermore National Laboratory, January 1988.

[6] R. G. BUSACKER AND T. L. SAATY, *Finite Graphs and Networks. An Introduction with Applications.* McGraw Hill Co, New York, NY. 1965.

[7] J. C. BUTCHER, *The Numerical Analysis of Ordinary Differential Equations.* John Wiley and Sons, New York, NY, 1987.

[8] J. C. CAVENDISH. M. J. WENNER. J. BURKARDT, C. A. HALL. AND W. C. RHEINBOLDT, *Dem: A new computational approach to sheet metal forming problems,* Int. J. for Num. Meth. in Eng.. 23 (1986), pp. 847–862.

[9] ———, *Punch stretching of sheet metal and differential equations on manifolds.* Int. J. for Num. Meth. in Eng., 25 (1988), pp. 269–282.

[10] L. O. CHUA AND P. M. LIN. *Computer-Aided Analysis of Electronic Circuits.* Prentice Hall, Englewood Cliffs, NJ, 1975.

[11] E. A. CODDINGTON AND N. LEVINSON, *Theory of Ordinary Differential Equations.* McGraw Hill Co, New York, NY, 1955.

[12] K. DEKKER AND J. G. VERWER, *Stability of Runge-Kutta Methods for Stiff Differential Equations,* North Holland, Amsterdam, The Netherlands, 1984.

[13] P. DEUFLHARD, E. HAIRER. AND J. ZUGCK, *One-step and extrapolation methods for differential-algebraic systems*. Numer. Math.. 51 (1987). pp. 501-516.

[14] P. DEUFLHARD AND U. NOWAK, *Extrapolation integrators for quasilinear implicit ODE*, in Large-Scale Scientific Computing, P. Deuflhard and B. Engquist. eds.. Birkhauser Verlag. Basel, Switzerland, 1987, pp. 37-50.

[15] R. FRANK, A. SCHNEID, AND C. W. UEBERHUBER. *Stability properties of implicit Runge-Kutta methods*, SIAM J. Numer. Anal., 22 (1985). pp. 515-534.

[16] F. R. GANTMACHER, *The Theory of Matrices, 2 Volumes*. Chelsea Publ. Co.. New York, NY, 1959.

[17] C. W. GEAR, *Simultaneous numerical solution of differential-algebraic equations*. IEEE Trans. Circuit Theory, CT-18 (1971), pp. 89-95.

[18] ——, *Differential-algebraic equation index transformations*, SIAM J. Sci. and Stat. Comp., 9 (1988), pp. 39-47.

[19] C. W. GEAR, H. H. HSU, AND L. E. PETZOLD, *Differential/algebraic systems revisited*. in Proc. of ODE Meeting, Mathem. Forschungsinst. Oberwolfach. Germany, 1982,

[20] C. W. GEAR, B. LEIMKUHLER, AND G. K. GUPTA. *Automatic integration of Euler-Lagrange equations with constraints*, J. Comp. Appl. Math., 12/13 (1985). pp. 77-90.

[21] C. W. GEAR AND L. R. PETZOLD, *Differential-algebraic systems and matrix pencils*, in Matrix Pencils, B. Kagstrom and A. Ruhe, eds., Lect. Notes in Math. Vol. 973, Springer Verlag, Heidelberg, Germany, 1983.

[22] E. GRIEPENTROG AND R. MAERZ, *Differential-Algebraic Equations and Their Numerical Treatment*, B.G.Teubner Verlag, Leipzig, Germany, 1986.

[23] E. J. HAUG, *Computer Aided Kinematics and Dynamics of Mechanical Systems. Vol. I: Basic Methods*, Allyn and Bacon, Boston, MA, 1989.

[24] E. J. HAUG AND J. YEN, *Implicit numerical integration of constrained equations of motion via generalized coordinate partitioning*, Tech. Report R-39, Univ. of Iowa, Ctr. for Simul. and Design Optim. of Mech. Systems, February 1989.

[25] A. C. HINDMARSH, *LSODE and LSODI, two new initial value ordinary differential equation solvers*, ACM-SIGNUM Newsletter, 15 (1983), pp. 10-11.

[26] J. W. HOOKER AND C. E.LANGENHOP. *On regular systems of linear differential equations with constant coefficients*, Rocky Mountain J. of Math.. 12 (1982). pp. 591-614

[27] P. LÖTSTEDT AND L. R. PETZOLD. *Numerical solution of nonlinear differential equations with algebraic constraints. Part 1: Convergence results for backward differentiation formulas*. Math. of Comp.. 46 (1986). pp. 491-516.

[28] N. K. MANI. E. J. HAUG, AND K. E. ATKINSON. *Application of singular value decomposition for analysis of mechanical systems dynamics*, Trans. ASME. 107 (1985). pp. 82-87.

[29] L. E. PETZOLD. *Differential/algebraic equations are not ODEs*, SIAM J. Sci. Stat. Comput.. 3 (1982). pp. 367-384.

[30] L. R. PETZOLD, *A description of DASSL: A differential/algebraic system solver*. in Scientific Computing, R. S. Stepleman, ed., North Holland, Amsterdam. The Netherlands. 1983, pp. 65-68.

[31] F. A. POTRA AND W. C. RHEINBOLDT, *On the numerical solution of Euler-Lagrange equations*, Mech. of Struct. and Machines, (1990) to appear .

[32] ——. *Differential-geometric techniques for solving differential algebraic equations*. in Real-Time Integration Methods for Mechanical Systems Simulation. NATO Advanced Workshop, 1989, E. Haug and R. Deyo, eds., Springer Verlag. Heidelberg. Germany, 1990.

[33] F. A. POTRA AND J. YEN, *Implicit numerical integration for Euler-Lagrange equations via tangent space parametrization*. Tech. Report R-56, Univ. of Iowa. Ctr. for Simul. and Design Optim. of Mech. Systems, June 1989.

[34] A. PROTHERO AND A. ROBINSON, *On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations*, Math. of Comp.. 28 (1974), pp. 145-162.

[35] P. J. RABIER, *Implicit differential equations near a singular point*, J. Math. Anal. and Appl., 144 (1989), pp. 425-449.

[36] P. J. RABIER AND W. C. RHEINBOLDT, *A general existence and uniqueness theory for implicit differential-algebraic equations*, Tech. Report ICMA-90-145, Univ. of Pittsburgh, Inst. for Comp. Math. and Appl., January 1990. Diff. and Int.Eq.. submitted.

[37] S. REICH. *On a geometrical interpretation of differential-algebraic equations.* Tech. Report. Techn. Univ. Dresden. Sekt. Inf. Technik, 1989.

[38] W. C. RHEINBOLDT. *Differential-algebraic systems as differential equations on manifolds,* Math. of Comp., 43 (1984), pp. 473–482.

[39] ——, *Numerical Analysis of Parametrized Nonlinear Equations.* John Wiley and Sons, New York, NY, 1986.

[40] ——. *On the computation of multi-dimensional solution manifolds of parametrized equations,* Numer. Math., 53 (1988), pp. 165–181.

[41] ——. *On the existence and uniqueness of solutions of nonlinear semi-implicit differential-algebraic-equations,* Tech. Report ICMA-90-139, Univ. of Pittsburgh, Inst. for Comp. Math. and Appl., August 1989, Nonlin. Anal. Theory and Appl., (1990) to appear.

[42] M. ROCHE. *Rosenbrock methods for differential-algebraic equations,* Numer. Math., 52 (1988), pp. 45–63.

[43] R. F. SINCOVEC, A. M. ERISMAN. E. L. YIP. AND M. A. EPTON. *Analysis of descriptor systems using numerical algorithms,* IEEE Trans.Aut.Control, AC-26 (1990), pp. 139–147.

[44] M. SPIVAK, *A Comprehensive Introduction to Differential Geometry. 5 Volumes.* Publish or Perish. Inc., Berkeley, CA, 1979.

[45] F. TAKENS, *Constrained Equations; A Study of Implicit Differential Equations and their Discontinuous Solutions,* Lect. Notes in Math. Vol. 525, Springer Verlag, Heidelberg, Germany, 1976, pp. 143–234.

[46] G. VERGHESE AND B. LEVY AND T. KAILATH, *A generalized state space for singular systems,* IEEE Trans. on Autom. Control, AC-26 (1981)

[47] N. M. WANG AND B. BUDIANSKY, *Analysis of sheet metal stamping by a finite-element method,* J. Appl. Mech., Trans. AMSE, 45 (1978).

[48] R. A. WEHAGE AND E. J. HAUG, *Generalized coordinate partitioning for dimension reduction in analysis of constrained dynamic systems,* Trans. ASME, 104 (1982), pp. 247–255.

[49] J. WITTENBURG, *Dynamics of Systems of Rigid Bodies,* B. G. Teubner, Stuttgart, Germany, 1977.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ICMA-90-149	**2. GOVT ACCESSION NO.**	**3. RECIPIENT'S CATALOG NUMBER**
4. TITLE (and Subtitle) On the Theory and Numerics of Differential-Algebraic Equations		**5. TYPE OF REPORT & PERIOD COVERED** Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) W. C. Rheinboldt		**8. CONTRACT OR GRANT NUMBER(s)** N-00014-90-J-1025
9. PERFORMING ORGANIZATION NAME AND ADDRESS Dept. of Mathematics and Statistics University of Pittsburgh Pittsburgh, PA 15260		**10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS**
11. CONTROLLING OFFICE NAME AND ADDRESS ONR		**12. REPORT DATE** May 1990
		13. NUMBER OF PAGES 40
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		**15. SECURITY CLASS. (of this report)** UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Differential-Algebraic Equations

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This is a written version of a series of invited lectures on differential-algebraic systems of equations (DAEs) at the IVth SERC Numerical Analysis Summer School of Lancaster Univ. In line with the aims of the meeting these notes introduce some typical applications and basic properties of DAEs and then present an overview of recent, new existence theories for such systems based on differential geometric considerations and on a numerical approach derived from these theories. In the presentation the stress is on general concepts, results and applications rather than on detailed proofs.